Journal of Electrical Engineering and Information Technologies, Vol. **8,** No. 1 pp. 29–38 (2023)
Article 205
Received: April 4, 2023
Accepted: May 6, 2023
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 004.738.5:004.2/.3
https://doi.org/10.51466/JEEIT2381205029j

Original scientific paper

ANTICIPATED DEVELOPMENTS IN CLOUD SERVICES WITH A FOCUS ON THE INFRASTRUCTURE-AS-A-SERVICE (IaaS) MODEL

Dimitar Jakovleski, Toni Janevski, Borislav Popovski

Faculty of Electrical Engineering and Information Technologies,
"Ss. Cyril and Methodius" University in Skopje,
Rugjer Bošković bb, P.O.Box 574, 1001 Skopje, Republic of North Macedonia
dimi.jakovleski@gmail.com

A b s t r a c t: Delivery of fundamental processing, network and storage resources to users has never been easier through IaaS, a cloud computing model that utilizes the internet in order to provide users with scalable infrastructure on-demand. However, the technology is constantly being developed in order to provide better QoS to the end consumers, so efforts are being made to optimize every part of the IaaS model to make it function even better. This paper provides an explanation of cloud services, cloud service models with special attention to the IaaS model, as well as predicting future areas of research and development in the IaaS model, namely, eliminating interruptions in services, resource management, security, Serverless Computing, Edge Computing and containerization. Ways to implement ML/AI algorithms for resource allocation, containerization, security and predictive maintenance are also discussed. At the end of the paper, a review is given of the business aspects of IaaS.

Key words; cloud Services; cloud service models; IaaS; cloud resources; ML/AI

ОЧЕКУВАНИ РАЗВОЈНИ ТРЕНДОВИ КАЈ СЕРВИСИ ВО ОБЛАК СО ФОКУС НА МОДЕЛОТ IaaS

А п с т р а к т: Обезбедувањето на основни процесирачки, мрежни и мемориски средства никогаш не било полесно одошто преку користење на IaaS, модел на сервис во облак кој го користи интернетот за да им понуди на корисниците скалабилна инфраструктура во секое време. Сепак, технологијата константно се развива со цел да се постигне подобар квалитет на сервис за крајните потрошувачи, па се прават истражувања за оптимизацијата на секој дел од моделот IaaS да функционира уште подобро. Трудот во продолжение дава објаснување на сервисите во облак, моделите на сервиси во облак со посебно внимание врз моделот IaaS, како и предвидувања за идните истражувачки области и развојот на моделот IaaS, имено отстранувањето на прекините кај сервисите, менаџирањето со ресурсите, сигурноста, Serverless Computing, Edge Computing, како и контејнеризацијата. Исто така се дискутира за различните начини на имплементација на алгоритмите ML/AI кај алокацијата на ресурсите, контејнеризацијата, сигурноста и превентивното одржување на инфраструктурата. При крајот на трудот се дава осврт на бизнис-аспектите на IaaS.

Клучни зборови: сервиси во облак; модели на сервиси во облак; IaaS; ресурси во облак; ML/AI

1. INTRODUCTION

With the development of the internet and computer software technology, there is an increasing demand for cheap processing power and memory, leading to a growing interest in using cloud services. Essentially, the cloud represents a concept of virtualized resources that are offered as a service over the internet. These resources can take the form of hardware, memory, networks, and software. "Renting" these resources has proven to be a fast and easy way to deploy applications and store data. Many companies have embraced this model and are transferring their infrastructure to the cloud, which, in addition to being financially viable, offers many other benefits, such as quick and easy infrastructure setup and

configuration, and access to data from anywhere at any time. Despite all the advantages, the lesser control over infrastructure and security threats can be considered as disadvantages of this way of deploying infrastructure in the cloud. However, work is underway to improve these already known disadvantages through various research that is already yielding results in making the cloud more secure and accessible to users. The use of the cloud gives organizations of all types and sizes the opportunity to grow faster and modernize their infrastructure. It has completely transformed the way we work, communicate and collaborate, and is becoming increasingly necessary for staying competitive in today's digital world. The Infrastructure-as-a-Service model offers greater freedom in managing resources in the cloud compared to other cloud service models, and as such, is of interest for research and future improvements to the model.

2. FUNDAMENTALS OF CLOUD SERVICES

Cloud services provide access to computer resources such as applications, servers (physical or virtual), data storage memory, development tools, network capabilities, and more, which are located in data centers maintained by cloud service providers. This "rental" of resources brings several advantages, including lower costs resulting from the elimination of the need to purchase, install, configure, and manage one's own infrastructure, greater agility in the infrastructure setup process, and easier and more cost-effective infrastructure scaling based on demand. Data suggests that by 2025, half of the spending on application software, infrastructure software and system infrastructure will be transferred to the cloud, which accounted for 41% in 2022 [1].

In general, cloud services consist of three main parts: cloud service providers store data and applications on physical machines at locations known as data centers, users use these resources, and the internet connects these two parts over long distances. Although these parts are simple, the technology that connects them is very complex. There are several types of cloud computing, namely public cloud, private cloud, hybrid cloud, and multi-cloud.

The public cloud is managed by cloud service providers who make it available to the public. They own all the hardware, software, and infrastructure that make up the cloud. However, their customers own the data and applications that reside on the cloud.

Private clouds are most often maintained by organizations such as corporations and universities to enable exclusive use.

Hybrid clouds combine the previous two types and take the best of both. The private part of the cloud is used for sensitive functions, while the public part is used to withstand increased demand for services from customers. This provides flexibility and security without having to abandon existing infrastructure and security [2].

Multi-cloud represents the use of multiple clouds from multiple different service providers. This offers the advantage of using each separate cloud based on the specific need at the moment.

The cloud operates through virtualization. With virtualization of the cloud, users can use only the parts of the services or resources they need, without the need to own physical infrastructure where the use of resources is not adequately planned. This means that users can quickly change the amount of required resources, which reduces costs and increases the flexibility of the system by leaving room for future expansions or reductions of required resources. Virtualization offers the possibility for the cloud service provider to virtualize their servers, disks, or other physical hardware, which in return offers a large number of services such as infrastructure, software, and platforms. Infrastructure-as-a-service provides users with access to cloud servers, disk space, and network resources. This means that customers do not need to buy their own infrastructure, but rather use the virtualized infrastructure in the cloud. Improving the use of resources is very important to enable a larger number of instances for the consumer. Data storage in the cloud is done on remote servers. These servers are maintained by service providers who are responsible for managing, hosting and securing the data stored in their infrastructure. The service provider is responsible for making the data on their servers always available to users through a public or private cloud. Through this storage method, capital costs are eliminated and a model with operational costs is adopted. Users input data into servers via the internet, where it is stored on virtual machines on physical servers. To ensure availability, data is distributed across multiple virtual machines in data centers located all over the world. If there is a greater need for memory, more virtual machines are simply created to meet the demand. Access to the data by the user occurs via the internet. Storing data in the cloud is beneficial for users who need backups of their data, wish to archive old data, or require the use of

powerful tools for processing and analyzing data that are provided.

Service-oriented architecture is a development and delivery practice that provides software as a series of interoperable services. The services are designed to be individual units with minimal interaction between them, with each service providing a part of the functionality. These individual services are then orchestrated to build an application that utilizes these services. Web services are a crucial part of cloud services. They are the most common way to access services in the cloud [3]. Figure 1 shows the division of Cloud Computing.

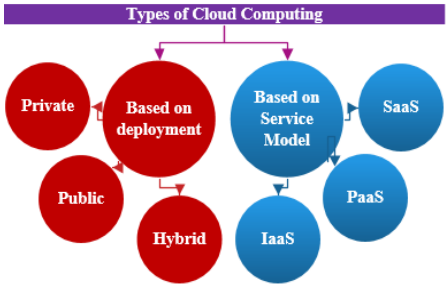

Fig. 1. Graphical representation of different types of Cloud Computing [4]

3. MODELS OF SERVICES IN THE CLOUD

The Software as a Service (SaaS) model provides software access through a subscription-based system, with the software located on external servers. Instead of each user having to install software on their own computer, they can access the program through the internet [5]. This model can be applied in several areas, including email services, product and service registration automation, document management, and more. Its advantages include enabling access from anywhere, lower costs, easy implementation, and easier upgrades, while its disadvantages include increased security threats associated with the way the model works, slower speeds, and less control. An example of SaaS is the Dropbox service, which is used as a cloud-based data storage service.

Platform as a Service (PaaS) is a cloud service platform that offers necessary software and hardware resources. This enables users to develop and manage their applications without maintaining the infrastructure required for software development. Essentially, PaaS provides an integrated solution to the user for application development, where code can be written, built, or managed without the need to install new versions of software and hardware, providing a complete development environment. This may include development frameworks, databases, and application servers [6]. A disadvantage is the limited control over the infrastructure. An example of PaaS is Red Hat OpenShift, which is a container management system that runs applications.

Infrastructure as a Service (IaaS) offers on-demand storage or virtualization in the cloud via the internet. Users are responsible for the operating system and any data and applications, but the service provider provides access to the network, servers, virtualization, and memory for data storage as needed. There is no need to maintain a proprietary data center, as the service provider handles all of that. Instead, the user gains access and control of the infrastructure through an API. IaaS offers flexibility in using only the necessary components and can be used as a fast way to set up and tear down development and testing environments. IaaS is commonly used in sales, as the number of sales during holiday periods typically increases, requiring a quick upgrade of infrastructure and the purchase of additional resources to handle the traffic during that period. The biggest drawbacks of this model are in the areas of security and trust, which can be overcome by selecting a good service provider that best meets the needs. Examples of IaaS include AWS, Microsoft Azure, and Google Cloud [7]. The difference between these cloud services can be seen in Table 1, by who is in charge of managing the resources.

T a b l e 1

Resources that are managed by either the user or a Service Provider (SP) depending on the cloud model

Model	IaaS	PaaS	SaaS
Applications	User	User	User
Data	User	User	SP
OS	User	SP	SP
Virtualization	SP	SP	SP
Servers	SP	SP	SP
Memory	SP	SP	SP
Network	SP	SP	SP

4. INFRASTRUCTURE-AS-A-SERVICE

IaaS, which stands for Infrastructure as a Service, is a model of cloud computing that provides access to virtualized resources via the internet. It is

one of three cloud service models, along with Platform as a Service (PaaS) and Software as a Service (SaaS). IaaS allows for quick scaling of required resources and helps users avoid the need for physical infrastructure. Each resource is offered as a separate component of the service. With this model, the service provider is responsible for the infrastructure, while the user installs and configures the software, including applications and operating systems. The main advantages of IaaS are:

• Users can utilize virtual resources without the need to purchase hardware.

• Infrastructure is scalable and can be easily changed based on the user's needs.

• It allows for virtualization of administrative tasks.

• The capital expense model is replaced with an operational expense model since the need to invest in hardware is eliminated.

Traditionally, the IaaS model has a problem with the pricing structure, which can be difficult to determine. Additionally, there is a lack of transparency from the service provider, making management and monitoring of systems challenging. Security is a significant issue as resources are accessed over the internet. Measures are being taken to improve the IaaS model to make it more attractive to users. Future trends indicate the implementation of various tools to improve resource management, control, security, and automation [7].

5. FUTURE TRENDS IN THE INFRASTRUCTURE-AS-A-SERVICE MODEL AND CLOUD SERVICES

The future of IaaS will likely depend on the growth and evolution of businesses that start to use the cloud instead of physical infrastructure. Users utilize IaaS to enable a wide range of functionalities, including software development, application testing, web hosting, and big data analysis. There are several obstacles and areas of development within the IaaS model that will define the future of cloud services.

a) *Elimination of service interruptions*

Interruptions in IaaS services are inevitable due to the nature of the model. The ultimate goal is to develop methods to minimize and even eliminate interruptions. These methods include quickly detecting the source of defects and resolving them. It is important to have some preparation in place in the systems and processes to enable a quick solution in case of a potential interruption. This problem represents one of the areas in which the IaaS model will develop in the future. There are several ways to build tolerance for interruptions in the infrastructure as a service, such as developing monitoring of the behavior and security of the components that manage access to the cloud infrastructure. Another way is to enable backup of data at multiple remote locations so that in case of a work interruption at one location, functionality can be provided by the remaining locations. This solution, however, represents an expensive alternative as it requires a significant investment in support infrastructure that would not be used every day.

b) *Resource menagement*

As the demand for IaaS increases, so will the demand for resources, making it of great importance for service providers to implement appropriate resource management models to achieve maximum profit. This need will drive interest in developing new, more efficient resource management models. The resources offered by the IaaS model should be allocated according to current needs, while ensuring good quality of service for end users. These resources are not unlimited, so it is important to distribute them efficiently. To effectively distribute resources, it is necessary to overcome obstacles such as resource provisioning, resource mapping, resource distribution, and resource adaptation.

1. *IaaS provisioning*

Provisioning of resources for users or applications through load balancers, mechanisms for enabling high availability, and the like is crucial for improving service quality in IaaS. In the future, there will be developments in the elasticity of applications hosted in the cloud in order to obtain an increased amount of resources and a reduced cost for their use, while keeping the parameters that the user requests in mind. Of interest also is progress in developing models for predicting resource needs in hosted applications in order to minimize the impact of increased resource usage with the least possible performance drop. Designing an algorithm for resource provisioning that converges to optimal processor allocation based on the rate of incoming data and processor power needs is yet another part of IaaS provisioning that is in development.

2. *Resource allocation*

When it comes to resource allocation, the optimal solution is achieved when the allocation made

to satisfy needs has the minimum possible cost for the infrastructure and in the shortest possible time period. Different resources are required for different needs of IaaS, so finding these resources and allocating them is of great importance. Allocation is performed according to predefined policies. Creating a resource allocation scheme that spans multiple clusters, creating mechanisms to control the cost-to-reconfiguration ratio and maximize cloud utilization, as well as relocating virtual machines while minimizing the system cooling consumption are just some of the areas of research when it comes to optimizing resource allocation.

3. *Resource mapping*

Resource mapping is one of the processes that build the system that enables the identification of existing resources and their application for specific needs. The problem here is achieving maximum utilization of the cloud in IaaS by calculating the capacity of the application's needs, in order to establish and maintain a minimal processing infrastructure. This is achieved through the use of a cognitive architecture that automatically creates a machine behavior model using previously collected data. Mapping logical nodes to physical nodes, finding physical resource allocation that corresponds to the needs of the logical network, developing models that predict how applications will perform based on various parameters such as processor, memory, network, and disk usage, as well as mapping application needs to cloud offerings to find a cloud service that best suits the specific application, are several areas that will be developed in the future.

4. *Resource adaptation*

Resource adaptation in the cloud is a crucial advantage, as it transitions to a model of operational costs and eliminates the need to purchase one's own infrastructure and hire personnel to maintain it, instead moving to a model where these resources are "rented". Such resources can easily be adjusted to everyday needs, avoiding over-provisioning. The most important part of resource adaptation is to maintain service quality and minimize costs, and progress in these areas is expected in the future [8].

c) *Security*

As the rate of migration to the cloud increases in the future, the motivation for cybercriminals to target cloud environments will also increase. Although cloud environments are protected by various security measures, they are still susceptible to security threats if not properly configured. In the future, attention will be focused on improving security, which will lead to the emergence of new security plans. Therefore, it is necessary to understand the development and security needs in the cloud to ensure a secure environment. In 2022, it was reported that 15 billion data records were stolen from 1.1 million compromised user profiles of 17 well-known companies, only through the method of "Credential Stuffing" [9].

1. *Cybersecurity mesh – CSMA*

The architecture of the network for cybersecurity is designed in a way that enables a scalable approach to extending security controls, even in distributed environments. It is built on a strategy of integrating distributed security tools through data centralization and control partitioning to achieve more effective collaboration among the tools. The cybersecurity network allows the tools to work together across multiple levels, such as consolidated policy management, security intelligence, and distributed identity frameworks that support identity and access management (IAM) frameworks. The result is improved detection capabilities, consistent policies, efficient responses, and adaptive access control. With this approach, a way is created for individual security services to communicate and integrate with each other, creating a dynamic security environment throughout the network. This protects all endpoints, rather than relying on a single technology to protect all assets. The use of this technology in IT creates the possibility of implementing a modular system that can be applied to multiple architectures simultaneously, centralizing security policy management. The advantage of this approach lies in its ease of implementation, practicality, and agility, making it a solution for the immediate needs brought about by digitalization, such as vulnerability to attacks, the cost of such attacks, and the growing migration to cloud-based systems. CSMA is also considered one of the building blocks of the Zero-Trust architecture.

There are four basic levels of CSMA:

• Security analysis and intelligence that enable the collection and analysis of large amounts of data from multiple locations at a single location, from which appropriate measures are taken.

• Security intelligence and distributed identity frameworks that support identity and access management frameworks, enabling adaptive access, decentralized identity management, and directing services.

• Consolidated policies that provide a central policy for all individual security tools, simplifying the detection of fraudulent attempts and problems.

• Consolidated control panels enable a more composite way of tracking security information obtained from systems, which helps with quick reactions and taking appropriate measures.

2. *Zero-trust architecture*

The Zero-Trust approach is a way to increase security through greater control of authentication, authorization, and security posture validation before allowing access to applications and data. This approach is modern and will be increasingly implemented in the future with the possibility of remote work, as it provides a way to give employees secure access to the applications and data they need from the work environment. By using Zero-Trust, attention is turned to security issues that arise as a result of hybrid cloud environments and hacker attacks. Implementation is done through a combination of advanced technologies such as multi-factor authentication, identity protection, user endpoint security, and system security maintenance. Traditional security approaches rely on end-users within an organization, which can lead to the risk of hacker attacks through the takeover of the credentials these end-users use to log into the system or the risk of internal attacks. The Zero-Trust architecture requires continuous monitoring and validation that end-users and their devices possess the appropriate privileges assigned to them. This approach is used in hybrid cloud systems, systems composed of multiple clouds, or multiple identities. It is used for SaaS applications and outdated systems. Zero-Trust adheres to the following basic principles:

• Continuous verification of all access attempts, protecting all existing resources.

• Minimization of potential damage from a possible attack by isolating users, giving them different privileges, or segregating the network to reduce the surface area that can be compromised by a security attack.

• Automating the processing of security data to accelerate the response process in the event of a security attack [10].

3. *Hybrid clouds and multi-cloud environments*

Full migration to a single cloud can be risky, so the model of deploying applications or data in different locations or clouds is increasingly chosen to ensure avoidance of complete losses. Hybrid cloud is a type of cloud service that combines private clouds (infrastructure physically located at the user's site) and public cloud. Hybrid clouds allow data and applications to be moved between the two environments. This model is used to comply with regulatory requirements, address latency issues, or maximize investments in proprietary infrastructure. Multi-cloud environments are a model in which cloud services from more than one cloud service provider are used in a heterogeneous environment. This allows for greater flexibility and reduces risk. Cloud services are chosen based on needs, so different service providers can offer services that are ideal for different applications. The use of this approach is growing and is expected to be one of the ways to ensure the existence of data and the operation of cloud-based applications [11].

4. *Security measures in the cloud*

The security approaches that are applied to the infrastructure owned by the user are not always sufficient for the cloud infrastructure, so special security measures are developed for the cloud. These measures enable the entire process of cloud applications, from infrastructure planning to deployment and maintenance, to be secured. One such security tool is Kube-bench, which checks whether the Kubernetes container orchestration tool is implemented according to best security practices. Some of the important measures include:

• Encryption of data, meaning that data is encrypted or ciphered during transmission and storage. There are various algorithms for implementing encryption, and the choice depends on the user's technical knowledge to prevent data loss.

• Network security, which involves dividing the network into multiple network segments, preventing external attacks, and allowing or denying access.

• Security checks through the use of security tools that serve to secure the infrastructure.

• Data backup in the event of unexpected loss due to natural disasters, which is often implemented by setting up backup servers at remote locations [12].

5. *DevSecOps*

DevSecOps is an approach that combines application development, security, operations, and IaaS in an automated CI/CD pipeline. The main goal

is to automate, monitor, and implement security in all phases of software development. Implementing security at every level of software development allows for cost reduction, reduced risk of security issues, and faster software deployment. The benefits of establishing a DevSecOps culture include increased quality and security of software, increased communication and collaboration between teams, faster resolution of security incidents, stronger security protocols, increased use of automation, particularly in the QA process, as well as earlier detection and correction of vulnerabilities in the code.

Practices used in implementing DevSecOps methods include:

• Implementation of automation to secure the CI/CD environment by adding security controls and testing throughout the entire development process.

• Dealing with open source software commonly used in application development. Implementation of automated processes that ensure code security is critical for this type of software tool.

• Adding security systems that provide information about the nature of defects, their severity, and the measures that need to be taken. This way, security risks can be resolved before they are deployed in production environments [13].

d) *Serverless computing*

This model allows service consumers to automatically allocate resources based on received user requests without the need for provisioning or server management by the user. In the context of IaaS, this model can be implemented through services such as AWS Lambda, Azure Functions, or Cloud Functions. These services enable the user to deploy their application and select a suitable function to process the application. The cloud service provider owns all the necessary infrastructure that automatically scales based on the user's needs, charging only for the resources used. The reduced need for infrastructure management, cost optimization through charging only for dynamically changing resources according to demand, scalability, and increased availability resulting from handling interruptions in functioning, motivate the development of serverless cloud services implementation [14].

e) *Edge computing*

By bringing processor and memory resources closer to data sources and users, latency and bandwidth requirements of traditional centralized implementations are reduced. Data is processed and analyzed at the edges of the network, where data is created or collected, rather than being sent to a centralized cloud for processing. This approach to resource management in IaaS environments is achieved by extending infrastructure to the network edges. This is done by creating virtual machines or containers on end devices such as routers, switches, and the like, which can "carry" applications and process data locally. This provides greater flexibility and scalability, and allows for filtering of data sent to the cloud [15].

f) *Containerization*

Containers represent a method of packaging and deploying applications in an easy and portable way, where each container contains all the necessary components to ensure the function of the application that resides on it. Containers can be applied in IaaS environments by deploying them on the virtual machines offered by the service provider. In this way, the advantages of containerization such as isolation, scalability, and portability are combined with the benefits offered by the IaaS model. In such an implementation, the user will be responsible for deploying the required container images and orchestrating the containers themselves. Containers in IaaS can simplify the deployment and management of applications by offering a consistent and portable environment. They can also offer scalability and optimal resource utilization by enabling multiple containers to run on a single virtual machine, as well as easy deployment of new versions of applications based on current needs. When it comes to container orchestration, platforms like Kubernetes or Docker Swarm can be used to further simplify the work and management of the containers on which applications are deployed.

6. ML/AI IN CLOUD SERVICES

AI and ML can be used to improve the function of the IaaS model by optimizing resource allocation and managing IaaS environments through analyzing past environmental behavior. They can provide assumptions about the future behavior of the system and its needs and automatically allocate resources to meet these needs. This improves performance, reduces costs, and improves overall resource utilization.

The use of the Markov Decision Process (MDP) and Bayesian learning helps for the optimization of dynamic resource allocation when implemented in cloud computing services. MDP helps in the alloca-

tion of resources for network function virtualization, while Bayesian learning aids in predicting the future utilization of resources by analyzing existing patterns. These models proved better than the traditional way of resource allocation in terms of cost [16]. The cost optimization of resource usage is a very important parameter in resource allocation. ML can be used in order to cut unnecessary costs from bad resource allocation strategies by identifying resources that are consistently underutilized and scaling them down or completely turning them off. In [17], a hierarchical framework is proposed that tends to solve both resource allocation and power management problems. It helps to perform the local power management of servers in an online and distributed way.

In terms of containerization and the different ways of container scheduling, the Kubernetes Container Scheduling Technique (KCSS) proves to improve the scheduling efficiency of containers. This is achieved by selecting the most suitable nodes together with a combination of dynamic needs from the customer and the status of the cloud computing for each requested container [18].

AI and ML also enhance security in IaaS environments by identifying potential threats and vulnerabilities. They learn from system data such as network traffic and analyze anomalies that could indicate security threats. This is achieved by finding patterns of behavior that can lead to security threats. ML algorithms can be trained to use these patterns in recognizing potential threats in network traffic like the different types of attacks, for example the Distributed Denial of Service (DDoS) attacks, phishing attempts and SQL injection attacks. Other implementations of AI and ML in IaaS can be found in access control and authentication mechanisms and automation of incident response in IaaS environments. Many models can be created for improving IaaS environment security in many different ways. One way is through supervised Artificial Neural Networks (ANNs), like the Levenberg-Marquardt (LMBP) algorithms, which besides improving security, can help in resource allocation, workload scheduling and energy optimization by minimizing the error between the predicted output and the actual output of the algorithm [19].

In this way, these security threats can be located and prevented before they occur, but it is important to mention that these algorithms can also become a target for attacks themselves, so it is best practice to combine them with other security measures. Also, like every other ML/AI algorithm, they rely on the accuracy of the models and a bad model can give incorrect data and generate false alarms, so it is important to use accurate models.

Predictive maintenance of infrastructure in IaaS through the use of ML and AL algorithms can help predict hardware component failures or other parts of the infrastructure. This reduces the time needed to service the infrastructure. The use of these tools can improve scalability by automating it, as well as automation in setting up loads, while minimizing downtime and making sure the infrastructure is working as intended. The Random Forest algorithm, as well as the Support Vector Machine (SVM) algorithm, for example, can be trained on data in order to predict infrastructure failures based on classifications of events.

7. BUSSINESS ASPECTS IN IaaS

IaaS opens up new horizons for businesses by providing the ability to transfer their entire infrastructure to the cloud. This migration offers complete elimination of capital expenditures (CapEx) and transition to an operational expenditure (OpEx) model. Additionally, the need for infrastructure maintenance, which would be maintained by the service provider in the cloud, is reduced. This way, attention is directed towards essential parts of the business that generate profit, such as product development, without worrying about the infrastructure on which the development occurs. This infrastructure setup, which is based on subscription to cloud service providers, can bring cost reduction, which can be shared in a short-term cost reduction and a long-term cost reduction. Both types are equally important for increasing the company's profit and reducing investment risk. In addition to these cost reductions, there is also an increase in business productivity by enabling greater efficiency of infrastructure usage by the user. This increases the value of the company itself. Projected increase of total revenue in the cloud is shown in Table 2.

Table 2

Projected increase of total revenue in the cloud [20]

Year	2022	2023	2024	2025
Total revenue ($B)	544	655	779	917

a) *Short-term cost reductions*

Choosing the IaaS model as a way to set up infrastructure in companies with existing infrastructure brings a reduction in short-term costs by utilizing and paying only for the necessary resources without owning too many unused resources. For new companies, it brings a benefit in reducing initial investments. By reducing initial investments, these companies can allocate part of the budget intended for purchasing and setting up infrastructure to other areas of the business, making the initial state of the company much better. Also, by outsourcing the maintenance of the infrastructure to someone else, the need for personnel in the company is reduced, thereby reducing the required initial budget for salaries.

b) *Long-term cost reductions*

Owning your own infrastructure is not a one-time investment, it has a lifecycle and over time new investments will be needed for repairs and upgrades, something that the user does not need to worry about when using IaaS, through which they can get additional resources with a few clicks without having to worry about the lifespan of the infrastructure. In addition, owning your own infrastructure also requires regular installations of new technologies and other improvements. When this happens, the infrastructure functions partially or does not function at all, reducing the productivity of the company. The space occupied by the infrastructure, i.e. the data centers, can be used for other purposes, or in the case of renting the space, the need to pay rent can be eliminated. The reduced need for personnel to maintain the equipment also has long-term benefits in reducing the costs of the company. By eliminating capital costs and moving to an operating cost model, the user is able to adjust resources according to current needs. As a result, unnecessary costs for unused resources are avoided, as only the resources used are charged for.

8. CONCLUSION

Infrastructure as a service (IaaS) is a cloud model that provides virtualized resources such as servers, memory, and network infrastructure over the internet. IaaS allows users to access these resources on-demand without the need to invest in or maintain their own infrastructure. Users can choose this model for web hosting, data storage, backup, or application development. Some benefits offered through the use of IaaS are scalability, flexibility, and cost savings. The model provides scaling of the amount of resources used as needed, where the user only pays for the resources used. However, there are also areas where improvements can be made, such as implementing tools to improve the model's function.

Eliminating interruptions in services is of particular importance because each interruption represents a loss of money for the service provider and a loss of resource usage opportunity for the user. Optimizing resource management enables resources to be allocated to more users, thereby increasing profits. Through the use of artificial intelligence and machine learning, containerization, serverless computing, and edge computing in the IaaS model, improvements in the model's operation, i.e. its automation and optimization, are offered. Security is the most critical element in enabling the normal function of the model and is one of the main subjects of research and progress in recent times. Since the cloud operates over the internet, the risk of security attacks is increased, so new and improved security measures are developed daily to make the cloud impervious to hacking attacks.

Infrastructure as a service is a powerful tool that enables users to access and use virtualized resources in a flexible and relatively inexpensive way, making it a popular choice for use, and the popularity itself brings with it future innovations.

REFERENCES

[1] Ranger, S. (2022): The Zdnet website, https://www.zdnet.com/article/what-is-cloud-computing-everything-you-need-to-know-about-the-cloud/, 25 February 2022.

[2] Rountree, D.; Castrillo, I. (2014): *The Basics of Cloud Computing – Understanding the Fundamentals of Cloud Computing in Theory and Practice*.

[3] Malhotra, L.; Agarwal D.; Jaiswal, A. (2014): Virtualization in cloud computing, *J Inform Tech Softw Eng*., **4**,136.

[4] Prajakta, P.; BasuMallick, C. (2022): The Spiceworks website, https://www.spiceworks.com/tech/cloud/articles/what-is-cloud-computing/, 9 February 2022.

[5] Rani Dimpi; Ranjan Kumar Rajiv (2014): A comparative study of SaaS, PaaS and IaaS in cloud computing, *International Journal of Advanced Research in Computer Science and Software Engineering*, pp. 158–161.

[6] Youssef, A. E. (July, 2012): Exploring cloud computing services and applications, *Journal of Emerging Trends in Computing and Information Sciences*, Vol. **3** (6), 838–847.

[7] Bhardwaj, Sushil; Leena Jain; Sandeep Jain (2010): Cloud computing: A study of Infrastructure as a Service (IaaS),

International Journal of Engineering and Information Technology **2** (1), 60–63.

[8] Manvi, S. S.; Krishna Shyam G. (2013): Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey. *Journal of Network and Computer Applications* **41**, pp 424–440, https://doi.org/10.1016/j.jnca.2013.10.004

[9] Indusface, The indusface website, March 30 2022.

[10] Mehraj Saima; Banday M. Tariq (2020): Establishing a zero trust strategy in cloud computing environment, *International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 22 – 24 2020. DOI:10.1109/ICCCI48352.2020.9104214

[11] Hong, J.; Dreibholz, T.; Schenkel, J.A.; Hu, J. A. (2019): *An Overview of Multi-cloud Computing*. DOI: 10.1007/978-3-030-15035-8_103

[12] Sridhar, S.; Smys S. (2016): A survey on cloud security issues and challenges with possible measures, *International Conference on Inventive Research in Engineering and Technology (ICIRST 2016)*.

[13] Myrbakken, H.; Colomo-Palacios, R. (2017): DevSecOps: A Multivocal Literature Review, Conference: *International Conference on Software Process Improvement and Capability Determination*. DOI:10.1007/978-3-319-67383-7_2

[14] McGrath, G.; Brenner, P.R. (2017): Serverless computing: design, implementation, and performance, *IEEE 37th International Conference on Distributed Computing Systems Workshops*, pp 405–410.

[15] Cao, K.; Liu, Y.; Meng, G.; Sun, Q. (2020): An overview on edge computing research, *IEEE Access*, Vol. **8**, pp. 85714–85728, DOI: 10.1109/ACCESS.2020.2991734.

[16] Javad Hassannataj Joloudari; Roohallah Alizadehsani; Issa Nodehi; Sanaz Mojrian; Fatemeh Fazl; Sahar Khanjani Shirkharkolaie; H M Dipu Kabir; Ru-San Tan; U Rajendra Acharya (March 2022): *Resource allocation optimization using artificial intelligence methods in various computing paradigms*: A Review. DOI:10.13140/RG.2.2.32857.39522

[17] Liu, N. et al. (2017): A hierarchical framework of cloud resource allocation and power management using deep reinforcement learning, In: *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA*, pp. 372–382. DOI: 10.1109/ICDCS.2017.123

[18] Jorge Luis Diaz (October 2021): Artificial intelligence-based Kubernetes container for scheduling nodes of energy composition.

[19] Umer Ahmed Butt; Muhammad Mehmood; Syed Bilal; Hussain Shah; Rashid Amin; M. Waqas Shaukat; Syed Mohsan Raza; Doug Young Suh; Md. Jalil Piran (2020): *A review of machine learning algorithms for cloud computing security*.

[20] Oleksandra, The Codica website, https://www.codica.com/blog/saas-paas-iass-choosing-the-best-cloud-computing-model/, 22 July 2022.