International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Analysis of Various Challenges of Big Data-as-a-Service (BDaaS) and Testing for its Research Aspects

Santosh Kumar Sharma[1], Dr. Ajay Pratap[1] and Dr. Harsh Dev[2]

Submitted: 28/04/2023 **Revised**: 27/06/2023 **Accepted**: 07/07/2023

Abstract: Big Data-as-a-Service is utilised in the current environment to manage and process the enormous amounts of data generated daily from many sources. Since the data is stored on a cloud platform, there is a chance that the system will malfunction and be attacked in many ways. To protect user data stored in the cloud from unwanted access, numerous studies have been undertaken and numerous systems developed. It is, however, very difficult to determine which problems Big Data-as-a-Service platform faces more frequently. We have gathered information on the cloud provider, how BDaaS solutions are used, the goal, implementation issues, and BDaaS challenges resolved. In this study, we looked at the overall design of BDaaS and examined the potential obstacles that.

Keywords: *Big Data-as-a-Service (BDaaS), BDaaS Architecture, BDaaS Challenges, Cloud Computing, Security & Privacy*

1. Introduction

Using big data Big Data and cloud computing are the two main technologies on which the service is built. Our data is accessible via the BDaaS cloud platform, and blockchain technology is used to regulate access to it. For the purpose of papers, we are presenting foundational concepts for both blockchain and BDaaS in this part.

Foundation of BDaaS (Big Data-as-a-Service)

Using big data Services that pertain to big data management and are provided in a cloud environment can all be grouped under the term "service" as a whole. Big data as a service, or BDaaS, is comparable to cloud computing's XaaS models, which are based on the SaaS, IaaS, and PaaS models. By encapsulating various data as a service, BDaaS, a brand-new service category, tackles the variations in data structure and descriptions. Now, consumers can use the CSPs' services in accordance with their needs for data storage, processing, and visualisation while paying appropriately[19]. It has gained a lot of popularity among consumers by assisting them in boosting productivity and reducing costs. Customers of BDaaS have access to three levels of abstraction: BDIaaS, BDPaaS, and BDASaaS. Here, BDPaaS combines the properties of data as a service and database as a service, and big data analytics is regarded as software as a service, which is the collection of different data analytics tools as a software [20], [21]. BDIaaS combines the elements of Storage and Computing as a Service. Some significant benefits include cost-effectiveness, speedy decision-making, improved data visualisation, effective data management, prompt query results, and data analytics. Amazon, IBM, EMC, Microsoft, Google, SAP, and Oracle are just a few of the many IT behemoths that have populated the BDaaS market space and are largely focused on offering big data storage and analysis services. One EMC utilised for data storage and analysis is Greenplum, which offers storage services and permits users to use Hadoop for BDA. Amazon Workspace provides independent BDA services, and Microsoft provides BDA services via the Windows Azure Marketplace.

Because BDaaS is a cloud-based service and the data is dispersed across numerous servers, users' top concerns in this environment are security, privacy, and access methods. Data manipulation, data theft, data loss, and denial of service are all highly typical security issues in cloud computing platforms because they result from a lack of trust and control over the data. The majority of cloud service providers do not give access control or data security guarantees.Therefore, it becomes crucial that all parties involved in BDaaS reevaluate the concepts of data security and access control. Big-Data-as-a-Service is used for processing and managing large volumes of data that are produced from many sources. The platform for cloud computing stores the collected data. It might therefore have a single point of failure, making it simple for attackers to target. To secure the system and defend the data stored on big data in a cloud environment, many academics are working in this field.

[1]*Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow-226028*
[2]*Department of Computer Science, Pranveer Singh Institute of Technology, Kanpur*
** Corresponding Author Email:santosh.sharma.ddn@gmail.com*

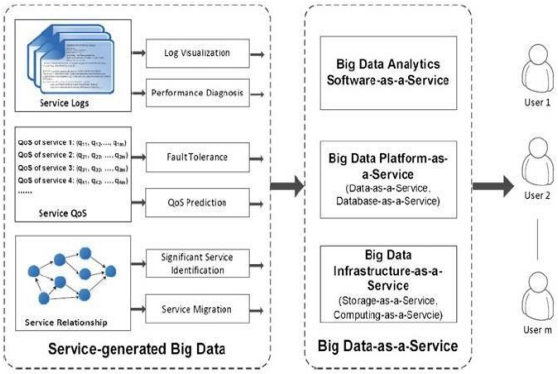

Fig 1. Architecture of Big Data-as-a-Service Platform

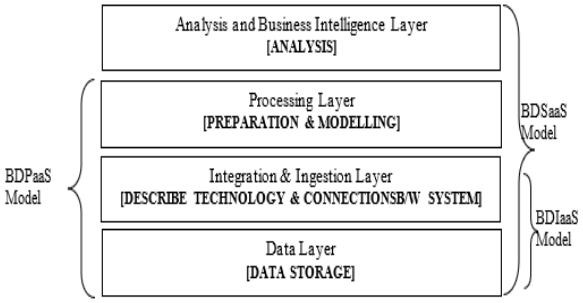

Fig 2. Formation of BDaaS Model

BDaaS Architecture

One of the newest study areas that combines big data with cloud computing is BDaaS. It protects against discrepancies that could arise because of data structure and definitions by encapsulating the data as a service. Users only care about what information or service they want to receive as well as when and where they can search, save, analyse, and visualise it. We are thinking about the BDaaS architecture that results from big data generated by services. Three main parts of service-generated big data—logs, service QoS, and service relationship—have been given by (Zheng et al. 2020). These parts are responsible for service identification, service migration, fault tolerance, fault log visualisation, and QoS prediction. These elements are the source of big data infrastructure as a service, big data platform as a service, and big data analytics software as a service. In this study, the BDaaS architecture for protected access control shown in Figure 1 is taken into consideration.

BDaaS Models and Its Challenges

We looked at three different Big Data as a Service (BDaaS) models in use today. These concepts closely resemble IaaS, PaaS, and SaaS, the three types of cloud infrastructure. The following are BDaaS models:We looked at three different Big Data as a Service (BDaaS) models in use today. These concepts closely resemble IaaS, PaaS, and SaaS, the three types of cloud infrastructure. The following are BDaaS models:

1. Basic data services, including as networks, virtual computers, and storage devices, are included in BDIaaS.
2. BDPaaS - It provides regularly used software that is offered by RedShift, EMR, or Amazon S3. Processes like ETL and BI are not included in this.
3. A full big data stack is provided by BDASaaS in a single tool.

2. Literature Survey

Due to the advancements being made in the field of computer sciences, there is a steady increase in the amount of data that is available and coming from various sources. However, users and consumers face significant challenges as a result of the abundance of data. The cloud computing environment is appropriate for storing and processing enormous volumes of data in the interim. Along with the characteristics of big data, [1] has offered a comparison of cloud service providers such Microsoft Azure, Google Cloud, Amazon Web Services, and International Business Machine cloud. Researchers have given an outline of some significant ideas and advancements in the field of supercomputing in their paper [2]. Additionally, this research paper discusses supercomputing-related current and next research questions.

Data security has emerged as one of the key security challenges that need more focus in the BDaaS environment. Huge data cloud computing principles, characteristics, and cutting-edge technologies are all covered in this research. Data quality and privacy management security issues connected to data access, data isolation, data integrity, deletion, transfer, and sharing are covered in depth [4]. The recent IT (Information Technology) boom that is greatly advancing IT departments across many enterprises is the cause of the combination of cloud and big data. This review [5] presents challenges with cloud computing that allow big data analytics. IoT devices are just meant to gather data and transport it for analysis, therefore data arriving from IoT devices has to be analysed. The writers have [6]

Cloud computing may be viewed as a benefit for any business that wishes to work on complex and large-scale computation since it provides a reliable infrastructure. In order to identify the difficulties of this setting, case studies of big data in the cloud environment have been provided in [8]. Due to the transformation in networks, IT platforms, and digital technologies, industry 4.0 has altered the factors determining organizations' innovation and competitiveness. Big data analytics' (BDA) future is more significant and will help firms that rely on data to perform better [9].

In this [10], writers explore the benefits of decentralising computing away from data centres as well as the evolving

cloud environment's infrastructure. The roadmap of issues that have been resolved to realise the benefits of next-generation cloud-based systems is also covered in the paper. The impacts of BDAC on the performance of any organisation were studied by the authors in [11]. Additionally, it analyses the connection between BDAC and FPER as well as the process of mediating the impacts of process-oriented dynamic capacities (PODC). Hadoop is used to study big data for land and ocean territory [13]. To explain how a design works, various computations are suggested to account for land and ocean ranges. [13]. The authors of this study [14] have suggested a columnwise high-order.

During the big data processing phase, data quality (DQ) is regarded as a crucial component. Experiments have been done in [15] to demonstrate the improved quality scores obtained from applying the newly found and improved DQRs to data. The purpose of the authors' work is to define the concept of big data and highlight the importance of big data analytics [16]. Understanding insights and information on the current big data platforms and their application areas, the benefits and drawbacks of big data tools, the application of big data analytics methodologies, and new research prospects is crucial for the development of big data systems in the future [18]. A summary of service-generated big data and big data-as-a-service has been given in order to solve these challenges. Three different sorts of service-generated big data are utilised to improve the performance of any system. Big Data platform-as-a-service, big data infrastructure-as-a-service, and big data analytics software-as-a-service are all parts of big data as a service. These might be used to provide consumers common big data-related services, improving efficiency and lowering costs [19].

3. Research Methodology

The goal of this study is to ascertain if various issues, such as those related to BDaaS service selection, BDaaS solution usage, effect of BDaaS solution use, and BDaaS solution implementation, have an influence on the business. It was discovered after completing a study of the relevant literature that there are still a few research on Bigdata as a Service that need to be done. As a result, we draw on a number of research that are still infrequently conducted on the subject. We will create a methodology to assist us assess the degree to which any security issue is hurting any firm.

1- A preliminary poll was conducted on five fundamental factors, with the first one being to determine which cloud provider, among private, secured, and hybrid BDaaS solutions, is most suited for enterprises. The second question is posed to determine which BDaaS solutions are more popular in which market. The third one is constructed to comprehend how the development of BDaaS is altering the analytics environment. The primary problem and difficulty addressed by the BDaaS paradigm are implementation-based difficulties, which are addressed in fourth and fifth parameter analyses.
2- The following possible dangers are being examined:
3- Choosing a BDaaS service presents challenges
4- Issues related to the use of the BDaaS Solution
5- Challenges resulting from the use of the BDaaS Solution
6- difficulties with implementation

Data Collection:

For this study, we employed questionnaires that were sent via Google Forms to professionals in the big data, cloud computing, and BDaaS platform industries as well as academics and researchers. Basically, we collected data using two methods: a literature review and a questionnaire.

Data collection based on Literature review:

The authors reviewed and analysed a variety of relevant research papers from online journal databases such Scopus, Web of Science, Science Direct, ACM Digital Library, Google Scholar, and SpringerLink, as well as a few books and online content sources, to gather data for this study. About 26 research publications that were published between 2013 and May 2022 are included in the study. Figure 2 depicts the system used to determine the final research articles.

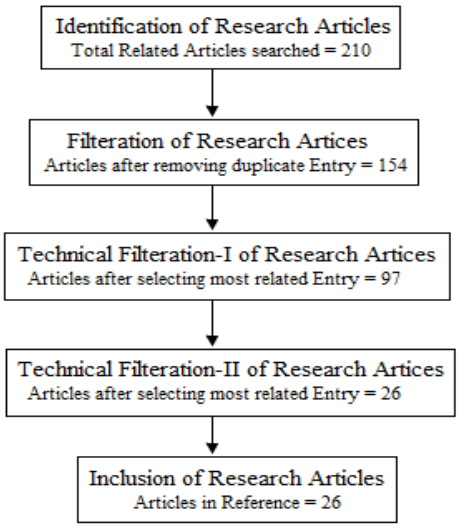

Fig 3: Selection of References

Data collection based on Questionnaire

To verify the research's hypotheses on the variables that affect the BSaaS platform's problems. The information was gathered through the dissemination of questionnaires created with Google form tools and given to relevant individuals. By splitting a group into equal halves and selecting random samples, the approach is used in sampling. The survey is set up using the Likert scale approach. A 96-person sample size was employed for data gathering, with 30 members of

industry, 36 academics, and 30 researchers in the same field serving as the study's population references.

Table 1.

SN	Variable	Symbol	Parameters
1	Cloud Provider	CP1	Private Cloud Provider
		CP2	Public Cloud Provider
		CP3	Hybrid Cloud Provider
2	Usage of BDaaS solutions	BS1	Customer Analytics (CRM)
		BS2	Performance Management
		BS3	Innovation
3	Analysis Objective of BDaaS	AO1	Affordability
		AO2	Minimal Use of Resources
		AO3	Real-Time Results
		AO4	Availability
4	Implementation Challenges for BDaaS	IC1	Lack of Preparation
		IC2	Crunch in Expertise
		IC3	Compromised Security
5	Challenges Handled by BDaaS	CH1	Data Quality and Diversity
		CH2	Data Security and Privacy
		CH3	Configurability

Table 2. Scale Chart

SN	Scale Category	Point
1	Strongly Agree	5
2	Agree	4
3	Neutral	3
4	Disagree	2
5	Strongly Disagree	1

Sampling and Setting of Hypothesis

Choosing the right sample size for a statistical study is among the most crucial responsibilities. Here, the confidence level has been set at 95%, with a z-score of 1.96, while the margin of error or confidence interval is 10%. The standard deviation is assumed to be 0.5 for the given population. We are taking into account 55,300,000 IT specialists worldwide. The sample size is determined using the formula below:

$$(n) = \frac{(zscore)^2 X\ SD\ X(1 - SD)}{e^2}$$

We obtain the sample size (n) = 96 by entering the values.

The claims that may be examined scientifically are known as hypotheses. Here, we're going to examine the correlation between the variables listed in Table 1.

Hypothesis 1:

H0, the null hypothesis The main problems that BDaaS attempts to solve are not those of data quality and diversity.

Alternative Hypothesis (H1): One of the main issues that BDaaS addresses is data quality and diversity.

Hypothesis-2:

Data Security and Privacy are not the main issues that BDaaS addresses, according to the null hypothesis (H0).

Data Security and Privacy are one of the primary difficulties tackled by BDaaS, according to the alternative hypothesis (H1).

Hypothesis-3:

The null hypothesis (H0) is that BDaaS does not handle configuration as a serious difficulty.

Alternate Hypothesis (H1): One of the main problems that BDaaS addresses is configurability.

4. Result and Discussion

We will assess the reliability and validity of the results from the survey. The reliability of the questionnaire has been determined using Cronbach's Alpha and the Average Variance Extracted (AVE) value, respectively. A trustworthy questionnaire is one where the estimated value is larger than 0.7.

Table 3. Validity Test Result

SN	Variable	AVE Value	Validity
1	Cloud Provider	0.921	Valid
2	Usage of BDaaS Solutions	0.924	Valid
3	Analysis of Objective of BDaaS	0.931	Valid
4	Implementation Challenges for BDaaS	0.931	Valid
5	Challenges Handled by BDaaS	0.933	Valid

Table 4. Reliability Test Result

SN	Variable	Cronbach Alpha Value	Reliability
1	Cloud Provider	0.922	Reliable
2	Usage of BDaaS Solutions	0.922	Reliable
3	Analysis of Objective of BDaaS	0.931	Reliable
4	Implementation Challenges for BDaaS	0.921	Reliable
5	Challenges Handled by BDaaS	0.931	Reliable

Table 5. T Test Result

SN	Variable	p-value	Result
1	Data Quality and Diversity	0.00009	alternative hypothesis: true
2	Data Security and Privacy	0.1995	alternative hypothesis: true
3	Configurability	0.00000	alternative hypothesis: true

We will test the hypothesis and compute the p value with a 5 percent threshold of significance after performing validity and reliability tests. All of the factors we considered have p-values that are less than 0.05, indicating a substantial impact. This indicates that the alternate hypotheses are substantial and that the null hypothesis for all of the hypotheses is invalid.

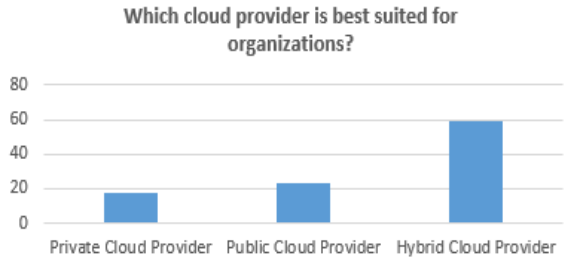

Fig 4: Selection of References

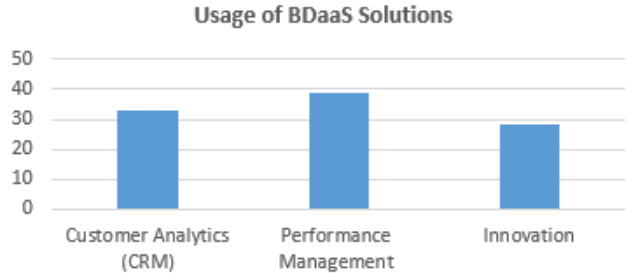

Fig 5: Selection of References

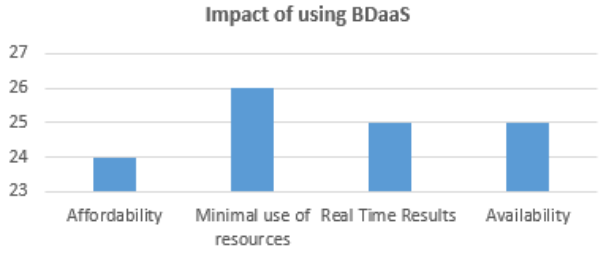

Fig 6: Selection of References

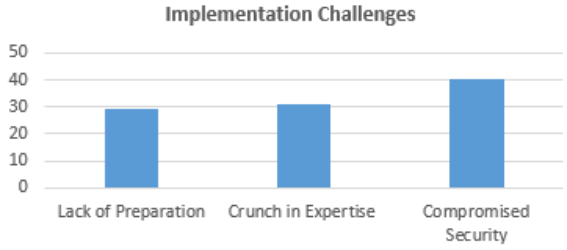

Fig 7: Selection of References

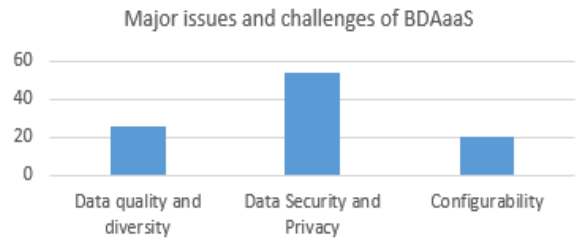

Fig 8: Selection of References

5. Conclusion:

The information in the result and discussion section leads us to the following deductions:

Organizations are best suited for hybrid cloud offerings. Businesses mostly utilise BDaaS systems for performance management, although consumer analytics and innovation may also be factors. The main benefit of employing BDaaS systems is the low resource use. Along with a lack of planning and a shortage of experience, compromised security can be one of the biggest implementation obstacles. In contrast to data quality, variety, and configurability, data security and privacy may be the main concern and problem of BDaaS. The conclusion of this research article points to several research areas in the fields of data security, data privacy, and access control.

Reference

[1] A. K. Sandhu, "Big data with cloud computing: Discussions and challenges," in *Big Data Mining and Analytics*, vol. 5, no. 1, pp. 32-40, March 2022, doi: 10.26599/BDMA.2021.9020016.

[2] Wei J, Chen M, Wang L, Ren P, Lei Y, Qu Y, Jiang Q, Dong X, Wu W, Wang Q, Zhang K. Status, Challenges and trends of data-intensive supercomputing. CCF Transactions on High Performance Computing. 2022. 1–20.

[3] Policy discussion – Challenges of big data and analytics-
driven demand-side management, Dr. Elizabeth Massey 1 (lead author), Dr. Piyush Verma 2, Mr. Beto Altamirano 3,
Mr. Andrei Covatariu 4, Mr. Romanas Savickas 5, https://unece.org/sites/default/files/2022-08/GEEE-9.2022.INF_.3-DataAnalytics_rev.pdf

[4] F. Wang, H. Wang and L. Xue, "Research on Data Security in Big Data Cloud Computing Environment," *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2021, pp. 1446-1450, doi: 10.1109/IAEAC50856.2021.9391025.

[5] N. S. Suhasini and S. Puli, "Big Data Analytics in Cloud Computing," *2021 Sixth International Conference on Image Information Processing (ICIIP)*, Shimla, India, 2021, pp. 320-325, doi: 10.1109/ICIIP53038.2021.9702705.

[6] A. Pratap, A. Kumar and M. Kumar, "Analyzing the need of edge computing for Internet of Things (IoT)", *Proc. 2nd Int. Conf. Comput. Commun. Cyber-Security*, pp. 203-212, 2021, DOI: 10.1007/978-981-16-0733-2_14

[7] Silva BN, Diyan M, Han K. Big data analytics. In: Khan M, Jan B, Farman H, editors. Deep learning: convergence to big data analytics. Singapore: Springer Singapore; 2019. p. 13–30.

[8] D. Tomar and P. Tomar, "Integration of Cloud Computing and Big Data Technology for Smart Generation," *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2018, pp. 1-6, doi: 10.1109/CONFLUENCE.2018.8443052.

[9] Vassakis K, Petrakis E, Kopanakis I. Big data analytics: applications, prospects and challenges. In: Skourletopoulos G, Mastorakis G, Mavromoustakis CX, Dobre C, Pallis E, editors. Mobile big data. Cham: Springer International Publishing; 2018. p. 3–20.

[10] Varghese, B. and Buyya, R. (2018). Next generation cloud computing: New trends and research directions. Future Generation Computer Systems, 79:849–861.

[11] Wamba SF, Gunasekaran A, Akter S, Ren SJF, Dubey R, Childe SJ. Big data analytics and firm performance: effects of dynamic capabilities. J Bus Res. 2017;70:356–65.

[12] Mehjabeen Sultana, "An Overview of Emerging Analytics in Big Data: In-Situ", International Journal of Computer Sciences and Engineering, Volume-04, Issue-05, Page No (166-169), May -2016, E-ISSN: 2347-2693.

[13] Tanuja A, Swetha Ramana D, "Processing and Analyzing Big data using Hadoop", International Journal of Computer Sciences and Engineering, Volume-04, Issue-04, Page No (91-94), Apr -2016, E-ISSN: 2347-2693

[14] X. Wang, W. Wang, L. T. Yang, S. Liao, and D. Yin, "A Distrib-1307 uted HOSVD method with its incremental computation for big1308 data in cyber-physical-social systems," IEEE Trans. Comput.1309 Social Syst., 2016.

[15] Taleb, I. and Serhani, M. A. (2017). Big data preprocessing: Closing the data quality enforcement loop. In Big Data (BigData Congress), 2017 IEEE

[16] P.Kodimalar, "A Study on Big Data and Big Data Analytical Research and Issues", International Journal of Computer Sciences and Engineering, Volume-03, Issue-11, Page No (171-179), Nov -2015, E-ISSN: 2347-2693

[17] Bernard Marr, "Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance", Wiley, 2015.

[18] Mohamed A, Najafabadi MK, Wah YB, Zaman EAK, Maskat R. The state of the art and taxonomy of big data analytics: view from new big data framework. Artif Intell Rev. 2020;53(2):989–1037.

[19] Zheng, Z., Zhu, J., and Lyu, M. R. (2013). Service generated big data and big data-as-a-service: an overview. In Big Data (Big Data Congress), 2013, IEEE International Congress on, pages 403–410. IEEE

[20] Tech Blogger Official Website, https://contenteratechspace.com/blogs/everything-you-need-to-know-about-big-data-as-a-service-bdaas/, accessed on 1 Dec 2022.

[21] Official Website of Tech Target, https://www.techtarget.com/searchcio/definition/big-data-as-a-service-bdaas, accessed on 12 Dec 2022.

[22] Official website of IGI Global, https://www.igi-global.com/dictionary/bdaas-big-data-platforms-as-a-service/71991, accesses on 15 Dec 2022.

[23] Official website of fortune b usiness insights, https://www.fortunebusinessinsights.com/big-data-as-a-service-bdaas-market-104738, accesses on 18 Dec 2022.

[24] Official website of semantix, https://semantix.ai/en/big-data-as-a-service-bdaas-como-funciona-e-quais-as-vantagens/, accessed on 2 Jan 2023.

[25] https://www.statista.com/statistics/1126677/it-employment-worldwide/

[26] Jaruwatcharaset, C. . (2023). Effects of Using a Temperature Control System in Bandicota indica Stalls with Internet of Things Technology. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 166–170. https://doi.org/10.17762/ijritcc.v11i4s.6324

[27] Prof. Madhuri Zambre. (2016). Automatic Vehicle Over speed Controlling System using Microcontroller Unit and ARCAD. International Journal of New Practices in Management and Engineering, 5(04), 01 - 05. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/47