

TOWARDS A RISK ASSESSMENT MODEL FOR BIG DATA IN CLOUD COMPUTING ENVIRONMENT

Saadia Drissi, Soukaina Elhasnaoui, Hajar Iguer,
Siham Benhadou and Hicham Medromi

(EAS-LRI) Systems Architecture Team, ENSEM, Hassan II University
Pluridisciplinary Laboratory of Research & Innovation (LPRI), EMSI
Casablanca, Morocco

ABSTRACT

Cloud computing gives a relevant and adaptable support for Big Data by the ease of use, access to resources, low cost use of resources, and the use of strong equipment to process big data. Cloud and big data center on developing the value of a business while reducing capital costs.

Big data and cloud computing, both favor companies and by cause of their benefit, the use of big data grows extremely in the cloud. With this serious increase, there are several emerging risk security concerns. Big data has more vulnerabilities with the comparison to classical database, as this database are stored in servers owned by the cloud provider. The various usage of data make safety-related big data in the cloud intolerable with the traditional security measures.

The security of big data in the cloud needs to be looked at and discussed. In this current paper, my colleagues and me will present and discuss the risk assessment of big-data applications in cloud computing environments and present some ideas for assessing these risks.

KEYWORDS

Cloud computing, risk assessment, big data, security

1. INTRODUCTION

In the recent years, there has been an acceleration use of big data in cloud computing environment, in several domains such as science, health, finance and government [2]. Cloud computing provides an important support for big data by accessing to resources, low cost and the usage of robust container to process big data. As the product big data needs to be stored in the cloud, the cloud or the container of data needs many servers linked to each other to distribute the computing tasks by using internet.

Cloud computing helps their users to use resources, access servers and store data based on demand. Cloud computing provides rapid dynamic and cheap computing power. Big data in cloud reduces the cost of operations since the companies do not have to buy their own product or service and to manage servers in same time by them self.

Big Data is somewhat dependent on the cloud for the flexibility that it provides. The processing of Big Data tools is then facilitated in an environment that can be adapted at will to optimize analytical operations. In fact, the union between Cloud Computing and Big Data is becoming a

good practice in the management of IT systems in many companies. The evolution of Big Data in cloud computing naturally raises the question of data protection and privacy respect. This is the biggest drawback of Big Data in cloud and the biggest challenge.

Big data and cloud computing, both favor companies and by cause of their benefit, the use of big data generates extremely significant positive cash flows in the cloud. With this serious increase, there are several emerging risk security concerns. Big data has more vulnerabilities with the comparison to classical database, as this database are stored in servers owned by the cloud provider. The various usage of data make protecting big data in the cloud intolerable with the traditional security measures [1].

In order to build a comprehensive risk assessment methodology for big data in cloud computing, a large literature review was conducted to identify all risk factors that may affect big data in cloud computing and major related research. Literatures on risk assessment for big data in the cloud are missing. Thus, additional effort must be employed in developing security risk assessment of big data in cloud computing.

The remainder of this paper is organized as follows: Section 2 provides a global overview of big data and cloud computing. Section 3 discusses and presents how big data and cloud computing work well together. Section 4 discusses the several risks to address in cloud computing and big data systems. Section 5 discuss risk assessment regarding big data in cloud computing and Section 7 presents the discussion, conclusions and future work.

2. FUNDAMENTAL CONCEPTS

Actually, the use of latest technology is essential for diverse IT operations and various industries such as big data, cloud computing, machine learning and IoT (Internet of things), for many of their applications for efficient management of the company, several concepts and definitions are discussed below [3].

2.1. Big Data

The concept of Big Data comes from a set of electronic operations from multiple data sources. This technology requires very important processing power and high capacities for data analysis and processing [7]. The importance of big data is focused in the analytical use of data, which helps to generate a clear decision to deliver better and faster services [8].

Big data is becoming a major innovation force in the field of research. This paradigm can be considered as a way to get and get appropriate information from large data set, providing information on huge data loads. As such, companies see this paradigm as a tool for understanding their customers, getting closer to them, finding patterns, and predicting trends. Additionally, scientists see big data as a way to store and process huge sets of scientific data. This concept is a hot topic for researchers and is expected to continue to gain popularity in the years to come.

The five different aspects used to describe big data are Volume, Variety, Velocity, Value and Veracity [4]:

- Volume is the amount of data coming from multiple sources, which show the huge data in numbers, this huge data can be measured in Gigabytes is now measured in Zettabytes or even Yottabytes. The volume is one of most evident dimension in big data characteristics.

- Variety is the data types, with the increase the number of Internet users in the world, data has changed from structured data in databases to unstructured data that contains a large number of formats such as images, audio and video clips, SMS, and GPS data [9].
- Velocity is the speed of data process and accessibility from several sources. The huge increase in data volume and their frequency requires the need for a system that ensures super-speed data analysis.
- Veracity is the quality of data; it shows the correctness of the data and the confidence in the data content. The quality of the data stored can change greatly, which influences the correctness of analysis. Although there is large agreement on the potential value of big data, the data is almost meritless if it is not perfect [10].
- Value is the value of big data, i.e. it indicates the importance of data after analysis. This is because the data on its own is almost valueless. The value lies in aware analysis of the precise data, the information and ideas it provides. The fifth characteristics of big data is the final stage that can be provided after processing volume, velocity, variety, contrast, validity and visualization [11]

2.2. Cloud Computing

Cloud computing is the realization of utility computing where the service provider implements resources and the cloud customer will pay as they use the resources. The user can access the cloud via a thin client. Cloud also provides memory for a large number of data to store and allows computation. Hence, many users can rely on a cloud as it reduces the infrastructure charge that the user needs to invest [5], [6].

Cloud providers offer three several basic services: Infrastructure as a Service (IaaS); Platform as a Service (PaaS); and Software as a Service (SaaS):

- **Software as a service (SAAS):** Cloud service providers offer different software applications to users who can use them directly without installation of software application on their computer. The user can adjust the settings and customizing the service as appropriate to his needs. SAAS helps big-data clients to perform data.
- **Platform as a service (PAAS):** Cloud service providers provide platforms, tools and other services to users, where the cloud service provider manages everything else, including the operating system and middleware, with resources that enable you to deliver everything from simple cloud-based apps to sophisticated.
- **Infrastructure as a service (IAAS):** Cloud service providers provide infrastructure such as storage, computing capacity, etc. is a form of cloud computing that provides virtualized computing resources over the Internet , In an IaaS model, a third-party provider hosts hardware, software, servers, storage and other infrastructure components on behalf of its users [12][13].
- **DaaS :** It is the alternative cloud computing model, as it differs from traditional models like (SAAS, IAAS, PAAS) in providing data to users through the network, as data is considered the value of this model [14] in conjunction with cloud computing based on solving some of the challenges in managing a huge amount of data. For these reasons, DaaS is closely related to big data whose technologies must be utilized [15]. DaaS provides highly efficient methods of data distribution and processing. DaaS is closely related to SaaS (storage as a service) and SaaS (software as a service) which can be combined with one of these models or both of them [16]. must appear as close to their point of reference as satisfactory formatting of the final document permits.

3. BIG DATA IN CLOUD COMPUTING

In [19], the usage of big data and cloud computing have been studied from several important aspects, and the authors have concluded that the relationship between them is complementary. Because of this big link between them, a model was prepared to show the relationship between them, and how much the both is compatible. Cloud computing can be considered as an environment of flexible distributed resources that uses high techniques to process and manage of data and yet reduces the cost. All these characteristics show that cloud computing has an integrated and consolidated relationship with big data. The twice are moving towards rapid progress to keep pace with progress in technology requirements and users. Big Data is somewhat dependent on the cloud flexibility. The processing of Big Data tools is then facilitated in an environment that can be adapted at will to optimize analytical operations.

Cloud computing and big data each go with the other, the first reason is that big data could support a huge data storage capacity in the cloud system, the second reason is that cloud computing uses different computing resources and storage for data analysis and processing. Thus, with a Big Data application having calculation capacities, Big Data progresses and accelerates the development of cloud computing. Distributed storage technology in environmental computing helps manage big data [17]. In [18], the authors claimed that the use of big data in the cloud makes data easily exposed. Cloud, which consists of many servers connected to each other.

Cloud and Big Data complement each other; cloud-based systems delivers high bandwidth, immense amounts of memory, and scalable processing power to help Big Data applications with enhanced real-time processing, storing and analyzing big volumes of data. Because of this corresponding complementary of these two technologies, we need to couple each one to others for our better technology's future. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant.

4. RISK ASSESSMENT REGARDING CLOUD COMPUTING AND BIG DATA

In the literatures, there are several research paper, and the authors discuss risk assessment from various perspectives. In [20], a risk evaluation model is proposed to solve the dynamic and fuzzy of security evaluation and to solve the problem of expert evaluation. In [21], a security risk assessment algorithm is claimed to predict security risk level. In [22], risk assessment metrics used to assess risk and continuous assessment techniques to ensure cloud service customers' trust on the security and privacy assurances of cloud service providers. In [23], risk assessment is discussed to processes obtained with the deployment of different security controls to provide automatic assessment of costs and risk factors. In [24], a systematic analysis of threats and vulnerabilities are introduced in risk assessment to provide a better security. In [25], a classification is used to segregate information based on the importance and level of protection required to protect privacy. Finally, in [26] the trust and control are discussed to reduce risk of cloud adoption.

While big data is a Swiss army knife that solves many of the current issues around high data volumes, it is an ever-evolving field that is still developing and still has some problems. In this section, we present some risks associated with Big Data.

Cloud computing has more than twenty biggest risks for digital enterprises.

Table I. Cloud Risk's Classification [27]

<i>Cloud risk's</i>	<i>Description</i>
Lock-in	Relying strongly on the services of one provider can lead to severe difficulties in changing the provider.
Loss of governance	When using Cloud services, the CC necessarily cedes control to the CP on a number of issues which may affect security
Supply chain failure	A CP can outsource parts of its production chain to third parties, or even use other CPs as part of its service. This way, a potential for cascading failures is created
Conflicts between customer hardening procedures and cloud environment	Certain security measures of a CC may conflict with a CP's environment, making their implementation by the CC impossible.
Social engineering attacks	Social engineering is understood to mean the art of manipulating people into performing actions or divulging confidential information.
Resource exhaustion (under or over provisioning)	As Cloud services are on-demand services, there is the possibility that the CP won't be able to meet an increased demand in a certain shared resource, or to maintain a given service level.
Isolation failure	In shared environments, errors or attacks can lead to situations where one tenant has access to another tenant's resources or data.
Cloud provider malicious insider - abuse of high privilege roles	Malicious insiders at the CP can cause various kinds of damage to a CC's assets.
Management interface compromise (manipulation, availability of infrastructure)	The customer management interfaces of public cloud providers are Internet accessible and mediate access to larger sets of resources
Intercepting data in transit	Whenever data is transferred between different computers or sites, there is the possibility that the transfer can be intercepted
Insecure or ineffective deletion of data	Deleting data from Cloud storage does not in fact mean that the data is removed from the storage or eventual backup media.
Distributed denial of service (DDoS)	Distributed Denial of Service attacks aim at overloading a resource flooding it with requests from many sources distributed across a wide geographical or topological area
Economic denial of service (EDoS)	As a consequence of attacks, poor budget planning, or misconfigurations, the cost of a Cloud service can strain the financial resources of a CC to an extent that the service is no longer affordable.
Compromise of Service Engine	The service engine is a fundamental part of a Cloud service. A compromise of the service engine will give an attacker access to the data of all customers
Loss of Cryptographic Keys	The loss or compromise of cryptographic keys used for encryption, authentication or digital signatures can lead to data loss, denial of services, or financial damages
Non Cloud-Specific Network-Related Technical Failures or Attacks	Cloud services can be affected by a number of network-related technical failures that can also occur on classic IT settings.
Loss of Backups	The backups a CP makes of it's customers' data can get lost, damaged, or the physical media on which the backup is stored can get stolen.
Natural disasters	Natural disasters like flooding, earthquakes, tsunamis can affect the infrastructure of a CP.
Subpoena and e-discovery	Law enforcement authorities may ask operators of IT infrastructures to provide information pertaining to criminal cases, or information may

<i>Cloud risk's</i>	<i>Description</i>
	have to be provided during civil lawsuits.
Risk from changes of jurisdiction	When data is stored or processed in a data centre located in a country other than the CC's.
Data protection risks	Processing data in another country may incur difficulties regarding data protection legislation
Licensing Issues	Violating a software supplier's licensing agreements can result in significant financial penalties or disruptions of service.
Intellectual Property Issues	Both in the Cloud and when using certain software and service environments within the own infrastructure.

There are the five biggest risks that big data presents for digital enterprises.

Table II. Big Data Risk's Classification[28]

<i>Big data risk's</i>	<i>Description</i>
Unorganized data	Big data is highly versatile. It comes from number of sources and in number of forms.
Data storage and retention	This is one of the most obvious risks associated with big data.
Cost management	The process of storing, archiving, analyzing, reporting and managing big data involves costs.
Incompetent analytics	Without proper analytics, big data is just a pile of trash lying unnecessarily in your organization.
Data privacy	With big data, comes the biggest risk of data privacy.

Literatures on risk assessment for big data in the cloud computing are lacking. Thus, we need to put more efforts on it so that we can adopt and use the big data even though they are hosting in the cloud environment.

5. SYNTHESIS AND DISCUSSION

In the literature mentioned below, there are no research paper that presents a clear and detailed solution or model in risk assessment for big data in cloud computing.

Risk assessment for big data in cloud computing is one of a primary role in safety-critical business, in order to implement the control measures necessary to ensure an acceptable level of safety industries. However, it faces a series of general challenges, in part related to technology development and increasing needs. The corresponding models, theories, technologies and methods of traditional risk assessment have been hard to deal with the huge amount of data and information generated by the advancement of modern technologies. New related technologies need further research and improvement.

There is currently a need for a dynamic risk assessment to identify and analyze all potential events that may negatively affect the business environment. Therefore, dynamic risk assessment relies on experience, training and continuing education and the definition of techniques to process relevant data, which must be used with adaptable capacity to deal with unforeseen events and provide the right support to enable risk assessment. Through this work, we propose a risk assessment model for Big Data in cloud computing based on machine learning, in particular a deep learning model.

The proposed model is a machine learning software that can assess huge amounts of data using the most appropriate methods in mathematics for company objective. This solution assesses all the probable risk factors as well as their features and analyse each of them depending on the probability of occurrence so that companies can rightly act to prevent any possible damage.

This model will help companies to deliver appropriate support for safety-related decision-making and propose the adapted security measures in an automatic and reactive way.

6. CONCLUSIONS

With the increase use of data on a daily base, big data systems have become one of major force of innovation that provides a way in managing information. Cloud environments strongly control big data solutions by providing adapted environments to big data systems.

Although big data in cloud computing is powerful systems that enable both enterprises and further research to develop and enhance, there are some concerns relating risk assessment that need further and real investigation and discussion. Additional effort must be employed in developing risk assessment security mechanisms for big data in cloud computing. Further research must be employed in near time to tackle this risk assessment security problem. Regarding this particular area, we are planning to use adaptable mechanisms in order to develop a solution for implementing a risk assessment model at several dimensions of big data systems running on cloud environments. In this current paper, we provided an overview of big data, cloud computing, big data in cloud environments, highlighting its complementary and its solid relationship and showing that both technologies work very well together but also presenting the challenges faced by the two technologies mainly in security risk assessment.

In the future work , we will present the risk assessment model for big data in cloud computing, the proposed model will based on deep learning to predict the different risks relating each environment and propose the adapted security measures in an automatic and reactive way.

REFERENCES

- [1] M. Paryasto, A. Alamsyah, and B. Rahardjo, "Big-data security management issues," 2014 2nd Int. Conf. Inf. Commun. Technol., pp. 59–63, 2014.
- [2] Hashem, I.A.T. et al., 2014. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47, pp.98–115.
- [3] S. Drissi, S. El hasnaoui, H. Iguer, S. Benhadou and H. Medromi., *Integration of Cloud Computing Big Data AI and Internet of Things: Review and open research issues*, DIGITECH2019 , At: Germany, 25/ 26 April 2019.
- [4] Sakr, S. & Gaber, M.M., 2014. *Large Scale and big data: Processing and Management* Auerbach, ed.,
- [5] S. Drissi, S. Benhadou and H. Medromi, "A new shared and comprehensive tool of cloud computing security assessment", In the *Proceedings of the Springer*, vol: 366 , *Advances in Ubiquitous Networking*, UNET 2015, Casablanca, Morocco, pp 155-167
- [6] S. Drissi, S Elhasnaoui, H Iguer, S. Benhadou and H. Medromi, " Security Risk Assessment of Multi-cloud System Adoption: Review and Open Research Issues", In the *Proceedings of the Springer*, *International Conference on Big Data and Smart Digital Casablanca*, Morocco, January 2019 DOI: 10.1007/978-3-030-12048-1_37
- [7] Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In *A decade in internet time: Symposium on the dynamics of the internet and society* (Vol. 21). Oxford: Oxford Internet Institute.
- [8] SHAN, Y. C., Chao, L. V., ZHANG, Q. Y., & TIAN, X. Y. (2017). Research on Mechanism of Early Warning of Health Management Based on Cloud Computing and Big Data. In *Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management 2016* (pp. 291-294). Atlantis Press, Paris.

- [9] Parvin Ahmadi Doval Amiri and Mina Rahbari Gavvani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. *Asian Journal of Information Technology*, 15: 2516-2525
- [10] Chen, Min, et al. *Big data: related technologies, challenges and future prospects*. Heidelberg: Springer, 2014.
- [11] Demchenko, Yuri, et al. "Big security for big data: Addressing security challenges for the big data infrastructure." *Workshop on Secure Data Management*. Springer, Cham, 2013.
- [12] Vacca, J. R. (Ed.). (2016). *Cloud Computing Security: Foundations and Challenges*. CRC Press. ch-15.
- [13] <https://support.rackspace.com/how-to/understandingthe-cloud-computing-stack-saas-paas-iaas/>
- [14] Terzo, O., Ruiiu, P., Bucci, E., & Xhafa, F. (2013, July). Data as a service (DaaS) for sharing and processing of large data collections in the cloud. In *Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on* (pp. 475-480). IEEE.
- [15] Motahari-Nezhad, H. R., Stephenson, B., & Singhal, S. (2009). Outsourcing business to cloud computing services: Opportunities and challenges. *IEEE Internet Computing*, 10(4), 1-17.
- [16] Rajesh Saturi, Data as a Service (Daas) in Cloud Computing [Data-As-A-Service in the Age of Data] Data as a Service Daas in Cloud Computing, *Global Journal of Computer Science and Technology Cloud & Distributed Volume 12 Issue 11*, 2012.
- [17] MALLICK, Pradeep Kumar (ed.). *Research Advances in the Integration of Big Data and Smart Computing*. IGI Global, 2015.
- [18] Hazirah Bee bt Yusof Ali, Lili Marziana bt Abdullah, *Systematic Literature Review of Risk Assessment for Big Data in Cloud Computing Environment: Security, Privacy and Trust*, AICCC '18, December 21–23, 2018, Tokyo, Japan.
- [19] Nabeel Zanoon, Abdullah Al-Haj, Sufian M Khwaldeh, *Cloud Computing and Big Data is there a Relation between the Two: A Study*, *International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017)* pp. 6970-6982
- [20] D. Zong-you, Z. Wen-long, S. Yan-an, and W. Hai-tao, "The application of cloud matter #x2014; Element in information security risk assessment," *2017 3rd Int. Conf. Inf. Manag.*, pp. 218–222, 2017.
- [21] S. Deng, D. Yue, X. Fu, and A. Zhou, "Security risk assessment of cyber physical power system based on rough set and gene expression programming," *IEEE/CAA J. Autom. Sin.*, vol. 2, no. 4, pp. 431–439, 2015.
- [22] R. Trapero, J. Luna, and N. Suri, "Quantifiably Trusting the Cloud: Putting Metrics to Work," *IEEE Secur. Priv.*, vol. 14, no. 3, pp. 73–77, 2016.
- [23] V. Bellandi, S. Cimato, E. Damiani, G. Gianini, and A. Zilli, "Toward economic-aware risk assessment on the cloud," *IEEE Secur. Priv.*, vol. 13, no. 6, pp. 30–37, 2015.
- [24] Lulu Liang, Wang Ren, Jing Song, Huaming Hu, Qiang He, and Shuo Fang, "The state of the art of risk assessment and management for information systems," *2013 9th Int. Conf. Inf. Assur. Secur.*, pp. 66–71, 2013.
- [25] V. Agrawal, "A Framework for the Information Classification in ISO 27005 Standard," *Proc. - 4th IEEE Int. Conf. Cyber Secur. Cloud Comput. CSCloud 2017 3rd IEEE Int. Conf. Scalable Smart Cloud, SSC 2017*, pp. 264–269, 2017.
- [26] A. Khosravani, "A case study analysis of risk, trust and control in cloud computing," *Conf. (SAI)*, 2013, pp. 879–887, 2013.
- [27] Catteddu, D., Hogben, G.: *Cloud Computing Information Assurance Framework*. ENISA (2009)
- [28] M. Paryasto, A. Alamsyah, and B. Rahardjo, "Big-data security management issues," *2014 2nd Int. Conf. Inf. Commun. Technol.*, pp. 59–63, 2014